

TALAGRAND’S TRANSPORTATION–COST INEQUALITY AND APPLICATIONS TO (ROUGH) PATH SPACES

SEBASTIAN RIEDEL

ABSTRACT. We give a new proof of Talagrand’s transportation cost inequality on Gaussian spaces. The proof combines the large deviation approach from Gozlan in [Goz09] with the Borell–Sudakov–Tsirelson inequality. Several applications are discussed. First, we show how to deduce transportation–cost inequalities for the law of diffusions driven by Gaussian processes both in the additive and the multiplicative noise case. In the multiplicative case, the equation is understood in rough paths sense and we use properties of the Itô–Lyons map to deduce the inequalities which improves existing results even in the Brownian motion case. Second, we present a general theorem which allows to derive Gaussian tail estimates for functionals on spaces on which a p -transportation cost inequality holds. In the Gaussian case, this result can be seen as a generalization of the “generalized Fernique theorem” from Friz and Oberhauser obtained in [FO10]. Applications to objects in rough path theory are given, such as solutions to rough differential equations and to a counting process studied by Cass, Litterer, Lyons in [CLL13].

INTRODUCTION

Transportation–cost inequalities can be seen as a functional approach to the concentration of measure phenomenon (see e.g. Ledoux’s work [Led01] for a survey on this topic). Let (E, d) be a metric space and let $P(E)$ denote the set of probability measures on the Borel sets of E . We say that the p -transportation–cost inequality holds for a measure $\mu \in P(E)$ if there is a constant C such that

$$(0.1) \quad \mathcal{W}_p(\nu, \mu) \leq \sqrt{CH(\nu | \mu)}$$

holds for all $\nu \in P(E)$. Here $\mathcal{W}_p(\nu, \mu)$ denotes the Wasserstein p -distance

$$\mathcal{W}_p(\nu, \mu) = \inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{E \times E} d(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}}$$

where $\Pi(\nu, \mu)$ is the set of all probability measures on the product space $E \times E$ with marginals ν resp. μ , and $H(\nu | \mu)$ is the relative entropy (or Kullback–Leibler divergence) of ν with respect to

Date: March 12, 2014.

2010 *Mathematics Subject Classification.* 28C20, 60F10, 60G15, 60H10.

Key words and phrases. concentration of measure, Gaussian processes, rough paths, stochastic differential equations, transportation inequalities.

Acknowledgements: SR is grateful for support from the European Research Council under the European Unions Seventh Framework Programme ERC grant agreement nr. 258237. Further, the author would like to thank Peter Friz and Michael Scheutzow for valuable discussions and comments.

μ , i.e.

$$H(\nu|\mu) = \begin{cases} \int \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise.} \end{cases}$$

If (0.1) holds, we will say that $T_p(C)$ holds for the measure μ .

Inequalities of type (0.1) were first considered by Marton (cf. [Mar86], [Mar96]). The cases “ $p = 1$ ” and “ $p = 2$ ” are of special interest. The 1-transportation-cost inequality, i.e. the weakest form of (0.1), is actually equivalent to Gaussian concentration as it was shown by Djellout, Guillin and Wu in [DGW04] (using preliminary results by Bobkov and Götze obtained in [BG99]). The 2-transportation-cost inequality was first proved by Talagrand for the Gaussian measure on \mathbb{R}^d in [Tal96] with the sharp constant $C = 2$ (for this reason it is also called *Talagrand’s transportation-cost inequality*). $T_2(C)$ is particularly interesting since it has the *dimension-free tensorization property*: If $T_2(C)$ holds for two measures μ_1 and μ_2 , it also holds for the product measure $\mu_1 \otimes \mu_2$ for the same constant C (see also [GL07] for a general account on tensorization properties for transportation-cost inequalities), and this property yields a dimension-free concentration of measure for μ . Gozlan realized in [Goz09] that also the converse is true: If μ possesses the dimension-free concentration of measure property, $T_2(C)$ holds for μ . We also remark that the 2-transportation-cost inequality gained much attention because it is intimately linked to other famous concentration inequalities, notably to the logarithmic Sobolev inequality: In their celebrated paper [OV00], Otto and Villani showed that in a smooth Riemannian setting, the logarithmic Sobolev inequality implies the 2-transportation-cost inequality. Since then, this result has been generalized in several directions, see e.g. the recent work of Gigli and Ledoux [GL13] and the references therein.

In this work, we will mainly study transportation-cost inequalities for the law of a continuous diffusion Y in a multidimensional setting, i.e. solutions to

$$(0.2) \quad Y_t = f_0(Y_t) dt + \sum_{i=1}^d f_i(Y_t) \circ dB_t^i; \quad Y_0 = \xi \in \mathbb{R}^m, \quad t \in [0, T]$$

where B is a d -dimensional Brownian motion, assuming that the vector fields $f = (f_i)_{i=0,1,\dots,d}$ are sufficiently smooth. In this context, $T_1(C)$ was first established with respect to the uniform metric by Djellout, Guillin and Wu in [DGW04]. Assuming that the solution Y is contracting in the L^2 sense (which implies the existence of a unique invariant probability measure), Wu and Zhang proved in [WZ04] that also $T_2(C)$ holds for the uniform metric. $T_2(C)$ is seen to hold for the weaker L^2 -metric also under milder assumption on the vector fields, cf. [DGW04] (see also [Wan02] in the context of Riemannian manifolds and [Sau12] where the Brownian motion B is replaced by a fractional Brownian motion (in the smooth setting with Hurst parameter $H > \frac{1}{2}$)).

A standard argument to establish transportation-cost inequalities, following [FÜ04] and [DGW04], is to use the Girsanov transformation. In the present paper, we introduce a new approach, where the key idea is to use Lyons’ *rough paths theory*. In the following, we will explain our strategy. The 2-transportation-cost inequality for Gaussian measures on Banach spaces reads as follows: Let (E, \mathcal{H}, γ) be a Gaussian Banach space, i.e. E is a Banach space, γ is a Gaussian measure and \mathcal{H} denotes the associated Cameron–Martin space. We set

$$(0.3) \quad d_{\mathcal{H}}(x, y) = \begin{cases} |x - y|_{\mathcal{H}} & \text{if } x - y \in \mathcal{H} \\ +\infty & \text{otherwise.} \end{cases}$$

The following Theorem was shown by Feyel and Üstünel, using the Girsanov transformation (cf. [FÜ04, Theorem 3.1]).

Theorem 0.1. *If (E, \mathcal{H}, γ) is a Gaussian Banach space and \mathcal{H} is densely embedded in E ,*

$$(0.4) \quad \inf_{\pi \in \Pi(\nu, \gamma)} \left(\int_{E \times E} d_{\mathcal{H}}(x, y)^2 d\pi(x, y) \right)^{\frac{1}{2}} \leq \sqrt{2H(\nu | \gamma)}$$

holds for all $\nu \in P(E)$.

Note that the inequality (0.4) does not really fit into the framework discussed above since $d_{\mathcal{H}}$ does in general not induce the topology on the space E . The statement of Theorem 0.1 is even more surprising since in the case $\dim \mathcal{H} = \infty$, we have $\gamma(\mathcal{H}) = 0$, in other words the function $d_{\mathcal{H}}$ equals $+\infty$ “very often”. The first contribution of the present work is to give a proof of Theorem 0.1 using the ideas of Gozlan [Goz09]; that is, we combine the dimension-free concentration property of the Gaussian measure γ (in the form of the Borell–Sudakov–Tsirelson inequality) with a large deviation argument. Theorem 0.1 applied to the Wiener measure (or more general Gaussian measures) on the space of continuous functions (or on a suitable subspace) will be our starting point. A fundamental observation made by Djellout, Guillin and Wu in [DGW04, Lemma 2.1] is that transportation-cost inequalities are stable under a push-forward by Lipschitz maps. If we could show that the map $I_f(\cdot, \xi)$, assigning to each Gaussian trajectory ω the solution trajectory $Y(\omega)$, is Lipschitz with respect to the metric $d_{\mathcal{H}}$, it would be immediate that $T_2(C)$ also holds for the law of Y . In the additive noise case, this is not hard to show and discussed in Section 2.1 for general Gaussian driving signals and different metrics. The multiplicative noise case is considerably more involved. It is well known that in this case, the map $I_f(\cdot, \xi)$ will in general not be continuous w.r.t. the uniform metric. However, the key result of Lyons rough paths theory in this context (cf. [Lyo98], [LCL07]) is that there is a metric space $\mathcal{D}_g^{0,p}$ and a measurable map S such that

$$\begin{array}{ccc} & \mathcal{D}_g^{0,p} & \\ & \uparrow S & \searrow \mathbf{I}_f(\cdot, \xi) \\ C_0 & \xrightarrow{I_f(\cdot, \xi)} & C_{\xi} \end{array}$$

commutes, where $\mathbf{I}_f(\cdot, \xi)$ is now a locally Lipschitz continuous function. Using this factorization, we can show a local Lipschitz continuity w.r.t. the metric $d_{\mathcal{H}}$. Recall the definition of the p -variation (pseudo-)metric $d_{p\text{-var}}$.

Proposition 0.2. *Let f be sufficiently smooth and choose $p \in (2, 3)$. Then there is a measurable function L such that*

$$d_{p\text{-var}}(y^1, y^2) \leq L(x^1) d_{\mathcal{H}}(x^1, x^2)$$

holds γ -almost surely for all $x^1, x^2 \in C_0([0, T]; \mathbb{R}^d)$ where γ denotes the Wiener measure and $y^i = I_f(x^i, \xi)$, $i = 1, 2$.

See Section 2.2 for a proof. The function L is very explicit, and it turns out that its $L^q(\gamma)$ moments are finite for every $q \in [1, \infty)$. A straightforward generalization of [DGW04, Lemma 2.1] leads to our main result in case of multiplicative noise.

Theorem 0.3. *Let μ denote the law of Y on the space $C_\xi([0, T], \mathbb{R}^m)$. Then for every $p \in (2, 3)$ and $\varepsilon > 0$ there is a constant C such that*

$$\inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{C_\xi \times C_\xi} d_{p\text{-var}}(x, y)^{2-\varepsilon} d\pi(x, y) \right)^{\frac{1}{2-\varepsilon}} \leq \sqrt{CH(\nu | \mu)}$$

for every $\nu \in P(C_\xi)$.

We make several remarks.

- If $x_0 = y_0$, $\|x - y\|_\infty \leq d_{p\text{-var}}(x, y)$ and therefore Theorem 0.3 also holds for the uniform metric.
- Rough paths theory allows to go beyond the usual semimartingale setting and we can consider more general Gaussian driving signals in (0.2) than Brownian motion. Saying this, Theorem 0.3 actually holds for much more general diffusions, namely for those which are driven by Gaussian rough paths in the sense of Friz–Victoir (cf. [FV10a]) and for which the Cameron–Martin space is continuously embedded in the space of paths with bounded variation. This is particularly the case for the fractional Brownian motion with Hurst parameter $H \geq 1/2$, and we extend results from Sausserau [Sau12] in the multidimensional setting by considering more general diffusion coefficients. See Section 2.2 for more examples.
- Already in the Brownian motion case, our results are interesting since we *almost* obtain the 2-transportation-cost inequality without the (rather strong) assumption that Y is contracting in L^2 (cf. [WZ04]). Even though we slightly fail to conclude the dimension-free concentration property, the forthcoming Theorem 0.4 shows why it is still desirable to obtain p -transportation-cost inequalities for $p \in (1, 2)$.

We finally discuss tail estimates for functionals on spaces on which a p -transportation-cost inequality holds. We cite a short form of Theorem 3.1 here.

Theorem 0.4. *Let V be a linear Polish space, $\mathcal{B} \subseteq V$ be a normed subspace and set*

$$d_{\mathcal{B}}(x, y) = \begin{cases} \|x - y\|_{\mathcal{B}} & \text{if } x - y \in \mathcal{B} \\ +\infty & \text{otherwise.} \end{cases}$$

Let $\mu \in P(V)$ and assume

$$\inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{V \times V} d_{\mathcal{B}}(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}} \leq \sqrt{H(\nu | \mu)} \quad \forall \nu \in P(V).$$

Let (E, d) be a metric space, $f: V \rightarrow (E, d)$ such that μ - a.s. for all $h \in \mathcal{B}$,

$$d(f(x + h), e) \leq c(x) (g(x) + \|h\|_{\mathcal{B}})$$

for some $e \in E$ with $c \in L^q(\mu)$, $\frac{1}{p} + \frac{1}{q} = 1$ and $\int c g d\mu < \infty$. Then $f: x \rightarrow d(f(x), e)$ has Gaussian tails.

Applied to Gaussian Banach spaces, Theorem 0.4 can be seen as a even more general form of the *generalized Fernique theorem* proved by Friz and Oberhauser in [FO10] (see also [DOR14]) where c had to be assumed to be bounded almost surely. Theorem 0.4 can be applied to many interesting examples arising from rough paths theory which are discussed more closely in Section 3.

The structure of the paper is as follows. Section 1 consists of a proof of Theorem 0.1. In Section 2, we establish transportation-cost inequalities for the law of diffusions; the additive noise case is treated in Section 2.1, the multiplicative noise case in Section 2.2. In Section 3 we consider

tail estimates for functionals on spaces in which transportation–cost inequalities hold and prove Theorem 0.4. In the appendix we collect some useful lemmata.

Notation. If (X, \mathcal{F}) is a measurable space, $P(X)$ denotes the set of all probability measures defined on \mathcal{F} . If X is a topological space, \mathcal{F} will be usually be the Borel σ -algebra $\mathcal{B}(X)$. If X and Y are measurable spaces and $\nu \in P(X)$, $\mu \in P(Y)$, then $\Pi(\nu, \mu)$ denotes the set of all product measures on $X \times Y$ with marginals ν resp. μ . If $c: X \times X \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ is measurable and $p > 0$, we set

$$\mathcal{W}_p^c(\nu, \mu) = \inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{X \times X} c(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}}$$

for $\nu, \mu \in P(X)$.

If $[S, T]$ is any interval in \mathbb{R} , we write $\mathcal{P}([S, T])$ for the set of all finite partitions of $[S, T]$ of the form $S = t_0 < t_1 < \dots < t_M = T$, $M \in \mathbb{N}$. If $x, y: [S, T] \rightarrow (B, \|\cdot\|)$ are paths with values in a normed space and $p \geq 1$, we define p -variation seminorm and pseudometric as

$$\|x\|_{p\text{-var}; [S, T]} := \sup_{D \in \mathcal{P}([S, T])} \left(\sum_{t_i \in D} \|x_{t_{i+1}} - x_{t_i}\|^p \right)^{\frac{1}{p}}; \quad d_{p\text{-var}; [S, T]}(x, y) := \|x - y\|_{p\text{-var}; [S, T]}.$$

If the time horizon is clear from the context, we sometimes omit the subindex $[S, T]$ in the notation. The set of all continuous paths $x: [S, T] \rightarrow B$ with $\|x\|_{p\text{-var}; [S, T]} < \infty$ is denoted by $C^{p\text{-var}}([S, T]; B)$ and we also define $C_\xi^{p\text{-var}}([S, T]; B) := \{x \in C^{p\text{-var}}([S, T]; B) : x_S = \xi\}$ for some $\xi \in B$. If B is a Banach space, $C_0^{p\text{-var}}([S, T]; B)$ is also a Banach space with the norm $\|\cdot\|_{p\text{-var}}$.

1. TRANSPORTATION INEQUALITY ON A GAUSSIAN SPACE

Let (F, \mathcal{H}, γ) be a Gaussian space where F is a separable Fréchet space, γ is a centered Gaussian measure on the Borel sets $\mathcal{B}(F)$ and \mathcal{H} denotes the corresponding Cameron–Martin space (cf. [Bog98] for the precise definitions). Recall that probability measures on Borel sets of Polish spaces are Radon measures, thus γ is Radon and \mathcal{H} is also separable (cf. [Bog98, 3.2.7. Theorem]).

In the next proposition, we prove a general form of $T_2(C)$ on Gaussian Fréchet spaces for continuous pseudometrics. The proof adapts ideas from [Goz09, Theorem 1.3], using the Gaussian free concentration property given by the Borell–Sudakov–Tsirelson inequality.

Proposition 1.1. *Let (F, \mathcal{H}, γ) be a separable Gaussian Fréchet space and let d_F be a metric which induces the topology on F . Let d be a pseudometric with the following properties:*

- (i) *There is a constant $L > 0$ such that $d(x, y) \leq L d_F(x, y)$ holds for all $x, y \in F$.*
- (ii) *There exists a constant C such that $d(x + h, x) \leq C|h|_{\mathcal{H}}$ for all $x \in F$ and $h \in \mathcal{H}$.*

Then

$$\inf_{\pi \in \Pi(\nu, \gamma)} \int_{F \times F} d(x, y)^2 d\pi(x, y) \leq 2C^2 H(\nu | \gamma)$$

holds for every $\nu \in P(F)$.

Remark 1.2. *Note the crucial fact that the constant L does not appear in the conclusion.*

Proof. We may assume w.l.o.g. that d is bounded (otherwise we prove the estimate for $d_n := d \wedge n$ instead and send $n \rightarrow \infty$ at the end, using Lemma 4.1). Thus, we may also assume that d_F is

bounded. Set

$$\mathcal{W}_2(\nu, \mu) := \inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{F \times F} d(x, y)^2 d\pi(x, y) \right)^{\frac{1}{2}}.$$

For $x = (x_1, \dots, x_n) \in X^n$, define

$$L_n^x := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}.$$

First, we claim that for every $x = (x_1, \dots, x_n) \in X^n$ and $h = (h_1, \dots, h_n) \in \mathcal{H}^n$,

$$(1.1) \quad |\mathcal{W}_2(L_n^{x+h}, \gamma) - \mathcal{W}_2(L_n^x, \gamma)| \leq C \frac{|h|_{\mathcal{H}^n}}{\sqrt{n}}.$$

Indeed: Since \mathcal{W}_2 is a pseudometric (cf. Lemma 4.2),

$$|\mathcal{W}_2(L_n^{x+h}, \gamma) - \mathcal{W}_2(L_n^x, \gamma)| \leq \mathcal{W}_2(L_n^{x+h}, L_n^x).$$

By the convexity property of $\mathcal{T}_2 := (\mathcal{W}_2)^2$ (cf. [Vil09, Theorem 4.8]) and assumption (ii),

$$\mathcal{T}_2(L_n^{x+h}, L_n^x) \leq \frac{1}{n} \sum_{i=1}^n \mathcal{T}_2(\delta_{x_i+h_i}, \delta_{x_i}) = \frac{1}{n} \sum_{i=1}^n d(x_i + h_i, x_i)^2 \leq C^2 \frac{|h|_{\mathcal{H}^n}^2}{n}$$

which shows (1.1). Now let $(X_i)_{i \in \mathbb{N}}$ be an i.i.d. sequence in F with law γ and let L_n be its empirical measure. Let m_n be the median of $\mathcal{W}_2(L_n, \gamma)$, i.e.

$$\mathbb{P}(\mathcal{W}_2(L_n, \gamma) \leq m_n) = \gamma^n \{x \in F^n \mid \mathcal{W}_2(L_n^x, \gamma) \leq m_n\} \geq \frac{1}{2}$$

and the same holds for the reversed inequalities. Define

$$A := \{x \in F^n \mid \mathcal{W}_2(L_n^x, \gamma) \leq m_n\}$$

and set

$$A^r := \{x + rh \mid x \in A, h \in \mathcal{K}^n\}$$

where \mathcal{K}^n denotes the unit ball in \mathcal{H}^n . If $x + h \in A^r$, (1.1) shows that

$$\mathcal{W}_2(L_n^{x+h}, \gamma) \leq \mathcal{W}_2(L_n^x, \gamma) + |\mathcal{W}_2(L_n^{x+h}, \gamma) - \mathcal{W}_2(L_n^x, \gamma)| \leq m_n + C \frac{r}{\sqrt{n}},$$

thus

$$\begin{aligned} A^r &\subset \left\{ x + h \mid \mathcal{W}_2(L_n^{x+h}, \gamma) \leq m_n + C \frac{r}{\sqrt{n}} \right\} \\ &\subset \left\{ x \in X^n \mid \mathcal{W}_2(L_n^x, \gamma) \leq m_n + C \frac{r}{\sqrt{n}} \right\}. \end{aligned}$$

Using the Borell–Sudakov–Tsirelson inequality (cf. [Bor75, Theorem 3.1]), we obtain

$$\begin{aligned} \mathbb{P} \left(\mathcal{W}_2(L_n, \gamma) > m_n + C \frac{r}{\sqrt{n}} \right) &= \gamma^n \left\{ x \in F^n \mid \mathcal{W}_2(L_n^x, \gamma) > m_n + C \frac{r}{\sqrt{n}} \right\} \\ &\leq \gamma_*^n((A^r)^c) \leq \bar{\Phi}(r) \end{aligned}$$

for all $r > 0$ where Φ denotes the cumulative distribution function of a standard normal random variable and $\bar{\Phi} = 1 - \Phi$. Equivalently,

$$\mathbb{P}(\mathcal{W}_2(L_n, \gamma) > u) \leq \bar{\Phi}\left(\frac{u - m_n}{C}\sqrt{n}\right)$$

for all $u > m_n$ and so

$$\frac{1}{n} \log \mathbb{P}(\mathcal{W}_2(L_n, \gamma) > u) \leq -\frac{\log(2)}{n} - \frac{1}{2} \left(\frac{u - m_n}{C}\right)^2$$

where we used the standard estimate $\bar{\Phi}(r) \leq 1/2 \exp(-r^2/2)$. From Varadarajan's Theorem (cf. [Dud89, Theorem 11.4.1]), with probability one, $L_n \rightarrow \gamma$ weakly in F for $n \rightarrow \infty$. Using [Vil03, Theorem 7.12] we see that $\mathcal{W}_2^{d_F}(L_n, \gamma) \rightarrow 0$ almost surely for $n \rightarrow \infty$. By assumption (i), also $\mathcal{W}_2(L_n, \gamma) \rightarrow 0$, hence $m_n \rightarrow 0$ for $n \rightarrow \infty$ and thus

$$(1.2) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathcal{W}_2(L_n, \gamma) > u) \leq -\frac{u^2}{2C^2}$$

for all $u > 0$. Note that for every $u > 0$, the set

$$\mathcal{O}_u = \{\nu \in P(F) \mid \mathcal{W}_2(\nu, \gamma) > u\}$$

is open in the weak topology. Indeed, assumption (i) implies that $\mathcal{W}_2 \leq L\mathcal{W}_2^{d_F}$, and since $\mathcal{W}_2^{d_F}$ metrizes weak convergence, $\nu \mapsto \mathcal{W}_2(\nu, \gamma)$ is continuous in the weak topology. Hence we may apply Sanov's Theorem (cf. e.g. [DZ98, Theorem 6.2.10]) which shows that

$$(1.3) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathcal{W}_2(L_n, \gamma) > u) \geq -\inf\{H(\nu \mid \gamma) \mid \nu \in P(F) \text{ such that } \mathcal{W}_2(\nu, \gamma) > u\}$$

and combining (1.2) and (1.3) we obtain

$$\inf\{H(\nu \mid \gamma) \mid \nu \in P(F) \text{ such that } \mathcal{W}_2(\nu, \gamma) > u\} \geq \frac{u^2}{2C^2}$$

which is the same as saying that

$$\frac{1}{2C^2} \mathcal{W}_2(\nu, \mu)^2 \leq H(\nu \mid \mu).$$

□

1.1. Banach spaces. In this section, we assume that $(B, \|\cdot\|)$ is a Gaussian Banach space and we set $d_B(x, y) := \|x - y\|$.

As an immediate corollary of Proposition 1.1 we obtain:

Corollary 1.3. *For any $\nu \in P(B)$,*

$$\inf_{\pi \in \Pi(\nu, \gamma)} \int_{B \times B} d_B(x, y)^2 d\pi(x, y) \leq 2\sigma^2 H(\nu \mid \gamma)$$

where

$$\sigma^2 = \sup_{l \in B^*, \|l\| \leq 1} \int l(x)^2 d\gamma(x) < \infty.$$

Proof. It is well known that $\sigma < \infty$ and that for every $h \in \mathcal{H}$ one has $\|h\| \leq \sigma|h|_{\mathcal{H}}$, cf. [Led96, Chapter 4], which gives the claim. □

Note that¹ the closure $\bar{\mathcal{H}}$ of \mathcal{H} in B coincides with the support of γ . Therefore, we may (and will) assume from now on that \mathcal{H} is continuously embedded B . Recall the definition of $d_{\mathcal{H}}$ given in (0.3). The key to proof our main theorem is the following Lemma which shows that the metric $d_{\mathcal{H}}$ can be approximated by metrics which fulfill the conditions of Proposition 1.1.

Lemma 1.4. *There is a sequence pseudometrics $(d_n)_{n \in \mathbb{N}}$ on B with the following properties:*

- (i) (d_n) is nondecreasing and $d_n \nearrow d_{\mathcal{H}}$ pointwise for $n \rightarrow \infty$.
- (ii) For every $n \in \mathbb{N}$ there is a constant L_n such that $d_n(x, y) \leq L_n \|x - y\|$ for all $x, y \in B$.
- (iii) $d_n(x + h, x) \leq |h|_{\mathcal{H}}$ for all $x \in B$, $h \in \mathcal{H}$ and $n \in \mathbb{N}$.
- (iv) All d_n are bounded.

Proof. Recall the following diagram:

$$B^* \hookrightarrow \mathcal{H}^* \hookleftarrow \mathcal{H} \hookrightarrow B.$$

Since the inclusion $i: \mathcal{H} \hookrightarrow B$ is dense, also $i^*: B^* \hookrightarrow \mathcal{H}^*$ is injective and has a dense image. This implies that we can find a complete orthonormal system $(e_n^*)_{n \in \mathbb{N}}$ in \mathcal{H}^* lying also in $i^*(B^*)$. If $R: \mathcal{H}^* \rightarrow \mathcal{H}$ denotes the Riesz identification map, the system $(e_n)_{n \in \mathbb{N}}$ defined as $e_n = R(e_n^*)$ is the dual system, i.e. $\langle e_m^*, e_n \rangle = \delta_{n,m}$. For such a system, we define maps $\pi_n: B \rightarrow \mathcal{H}$ as

$$\pi_n(x) = \sum_{k=1}^n \langle e_k^*, x \rangle e_k.$$

Note that π_n extends the orthonormal projection from \mathcal{H} onto the n -dimensional subspace $\mathcal{H}^n = \text{span}\langle e_1, \dots, e_n \rangle$, i.e. for $h \in \mathcal{H}$ we have

$$\pi_n(h) = \sum_{k=1}^n \langle e_k, h \rangle_{\mathcal{H}} e_k.$$

It follows that for $h \in \mathcal{H}$,

$$\pi_n(h) \rightarrow h$$

in \mathcal{H} for $n \rightarrow \infty$.

Set

$$\begin{aligned} \tilde{d}_n(x, y) &:= |\pi_n(x) - \pi_n(y)|_{\mathcal{H}} \quad \text{and} \\ d_n(x, y) &:= \tilde{d}_n(x, y) \wedge n. \end{aligned}$$

Clearly, all d_n are bounded pseudometrics. If $x, y \in B$,

$$\tilde{d}_n(x, y)^2 = \sum_{i=1}^n |\langle e_i^*, x - y \rangle|^2 \leq \sum_{i=1}^{n+1} |\langle e_i^*, x - y \rangle|^2 = \tilde{d}_{n+1}(x, y)^2$$

which shows that (d_n) is nondecreasing. Furthermore, if $x - y \in \mathcal{H}$,

$$\lim_{n \rightarrow \infty} d_n(x, y)^2 = \sum_{i=1}^{\infty} |\langle e_i, x - y \rangle_{\mathcal{H}}|^2 = |x - y|_{\mathcal{H}}^2 = d_{\mathcal{H}}(x, y)^2$$

¹This even holds more generally for Gaussian Radon measures on locally convex spaces, cf. [Bog98, 3.6.1. Theorem].

by Parseval's identity. Conversely, if $\lim_{n \rightarrow \infty} d_n(x, y) < \infty$ for some $x, y \in B$, we may define

$$\sum_{i=1}^{\infty} \langle e_i^*, x - y \rangle e_i =: z \in \mathcal{H}.$$

This implies that

$$\sum_{i=1}^{\infty} \langle e_i^*, x - y \rangle e_i = \sum_{i=1}^{\infty} \langle e_i^*, z \rangle e_i$$

and applying e_k^* on both sides shows that $\langle e_k^*, x - y \rangle = \langle e_k^*, z \rangle$ holds for every $k \in \mathbb{N}$. Hence $x - y = z \in \mathcal{H}$ and we have shown property (i). If $x, y \in B$,

$$\tilde{d}_n(x, y)^2 = \sum_{i=1}^n |\langle e_i^*, x - y \rangle|^2 \leq \|x - y\|^2 \sum_{i=1}^n \|e_i^*\|^2$$

which shows property (ii). For $x \in B$ and $h \in \mathcal{H}$,

$$\tilde{d}_n(x + h, x)^2 \leq \sum_{i=1}^{\infty} |\langle e_i, h \rangle_{\mathcal{H}}|^2 = |h|_{\mathcal{H}}^2$$

which finally gives property (iii). □

Now we can prove our main theorem.

Theorem 1.5. *For any $\nu \in P(B)$,*

$$\inf_{\pi \in \Pi(\nu, \gamma)} \int_{B \times B} d_{\mathcal{H}}(x, y)^2 d\pi(x, y) \leq 2 H(\nu | \gamma).$$

Proof. Take (d_n) as in Lemma 1.4. From Proposition 1.1,

$$\inf_{\pi \in \Pi(\nu, \gamma)} \int_{B \times B} d_n(x, y)^2 d\pi(x, y) \leq 2 H(\nu | \gamma)$$

holds for every $\nu \in P(B)$ and $n \in \mathbb{N}$. Sending $n \rightarrow \infty$ and using Lemma 4.1 shows the claim. □

1.2. Rough paths spaces. In the case of $B = C_0([0, T], \mathbb{R}^d)$, Theorem 1.5 immediately generalizes to rough paths spaces. Let γ be a Gaussian measure on B with corresponding Cameron–Martin space \mathcal{H} . For the sake of simplicity, we will assume that \mathcal{H} is continuously embedded in C_0 , otherwise we could have used a smaller space lying in C_0 instead. Let \mathcal{D} be a rough paths space (which could either be geometric or non-geometric, a p -variation or an α -Hölder rough paths space, cf. [LCL07], [FV10b] or [FH] for a precise definition) and assume that there is a measurable map $S: C_0 \rightarrow \mathcal{D}$ such that $\pi_1 \circ S = \text{Id}_{C_0}$ holds where $\pi_1: \mathcal{D} \rightarrow C_0$ is the projection map. The map S is called a *lift map*. Set $\gamma = \gamma \circ S^{-1}$. Abusing notation, we define $d_{\mathcal{H}}: \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R} \cup \{+\infty\}$ as

$$d_{\mathcal{H}}(\mathbf{x}, \mathbf{y}) = \begin{cases} |\pi_1(\mathbf{x}) - \pi_1(\mathbf{y})|_{\mathcal{H}} & \text{if } \pi_1(\mathbf{x}) - \pi_1(\mathbf{y}) \in \mathcal{H} \\ +\infty & \text{otherwise.} \end{cases}$$

Corollary 1.6. *For any $\nu \in P(\mathcal{D})$,*

$$\inf_{\pi \in \Pi(\nu, \gamma)} \int_{\mathcal{D} \times \mathcal{D}} d_{\mathcal{H}}(\mathbf{x}, \mathbf{y})^2 d\pi(\mathbf{x}, \mathbf{y}) \leq 2 H(\nu | \gamma).$$

Proof. By definition, $d_{\mathcal{H}}(S(x), S(y)) = d_{\mathcal{H}}(x, y)$, hence S is (in particular) 1-Lipschitz and the result follows from Theorem 1.5 and Lemma 4.3. \square

2. APPLICATIONS TO DIFFUSIONS

2.1. SDEs with additive noise. In this section we will consider SDEs with additive noise. Let $Y: [0, T] \rightarrow \mathbb{R}^d$ be the solution to

$$(2.1) \quad dY_t = dX_t + b(Y_t) dt, \quad Y_0 = \xi \in \mathbb{R}^d$$

where $X: [0, T] \rightarrow \mathbb{R}^d$ is a d -valued Gaussian process with continuous sample paths and $b \in C^1(\mathbb{R}^d, \mathbb{R}^d)$. It is well known that (2.1) has a unique solution given by $Y_t = I_b(X_t, \xi)$ where $I_b(\cdot, \xi): C_0([0, T]; \mathbb{R}^d) \rightarrow C_\xi([0, T]; \mathbb{R}^d)$ and $I_b(x, \xi) = y$ is defined as the unique solution to

$$(2.2) \quad y(t) = \xi + x(t) + \int_0^t b(y(s)) ds.$$

Lemma 2.1. *Assume that b is Lipschitz continuous with Lipschitz constant $L > 0$. Then for every $q \geq 1$,*

$$\|I_b(x + h, \xi) - I_b(x, \xi)\|_{q\text{-var}} \leq e^{LT} \|h\|_{q\text{-var}}$$

for every $x \in C_0([0, T], \mathbb{R}^d)$ and $h \in C_0^{q\text{-var}}([0, T], \mathbb{R}^d)$.

Proof. Set $y = I_b(x, \xi)$ and $y_h = I_b(x + h, \xi)$. Let $t_j \leq t_{j+1}$. Then,

$$\begin{aligned} |y_h(t_{j+1}) - y_h(t_j) - (y(t_{j+1}) - y(t_j))| &\leq |h(t_{j+1}) - h(t_j)| + \left| \int_{t_j}^{t_{j+1}} b(y_h(s)) - b(y(s)) ds \right| \\ &\leq |h(t_{j+1}) - h(t_j)| + L \int_{t_j}^{t_{j+1}} |y_h(s) - y(s)| ds. \end{aligned}$$

Hence

$$\begin{aligned} \|y_h - y\|_{q\text{-var}; [0, t]} &\leq \|h\|_{q\text{-var}; [0, t]} + L \int_0^t |y_h(s) - y(s)| ds \\ &\leq \|h\|_{q\text{-var}; [0, t]} + L \int_0^t \|y_h - y\|_{q\text{-var}; [0, s]} ds \end{aligned}$$

for every $t \in [0, T]$. The first estimate shows in particular that the left hand side is finite, therefore also the right hand side after the second estimate is finite. Applying Gronwall's Lemma shows that

$$\|y_h - y\|_{q\text{-var}} \leq \|h\|_{q\text{-var}} e^{LT}.$$

\square

Recall the definition of (fractional) Sobolev spaces: If $h: [0, T] \rightarrow \mathbb{R}^d$, set

$$|h|_{W^{\delta, p}; [s, t]} := \left(\iint_{[s, t]^2} \frac{|h(v) - h(u)|^p}{|v - u|^{1 + \delta p}} du dv \right)^{\frac{1}{p}}$$

for $\delta \in (0, 1)$, $p \in (1, \infty)$ and

$$|h|_{W^{1, p}; [s, t]} := \left(\int_s^t |\dot{h}(s)|^p ds \right)^{\frac{1}{p}}$$

for $p \in (1, \infty)$ where \dot{h} denotes the weak derivative of h . Set $\|h\|_{W^{\delta,p}} := \|h\|_{L^p,[0,T]} + |h|_{W^{\delta,p};[0,T]}$ and $d_{W^{\delta,p}}(x, y) := \|x - y\|_{W^{\delta,p}}$. The space $W^{\delta,p}$ consists of all paths h for which $\|h\|_{W^{\delta,p}} < \infty$ and $W_0^{\delta,p} = \{h \in W^{\delta,p} : h(0) = 0\}$. Both spaces are Banach spaces.

Lemma 2.2. *Assume that b is Lipschitz continuous with Lipschitz constant $L > 0$. Then for every $\delta \in (0, 1]$ and $p \in (1/\delta, \infty)$, there is a constant $C = C(\delta, p, L)$ such that*

$$\|I_b(x + h, \xi) - I_b(x, \xi)\|_{W^{\delta,p}} \leq C\|h\|_{W^{\delta,p}}$$

for every $x \in C_0([0, T], \mathbb{R}^d)$ and $h \in W^{\delta,p}$.

Proof. As before, set $y = I_b(\gamma, \xi)$ and $y_h = I_b(\gamma + h, \xi)$. It is easy to see that

$$(2.3) \quad \|y_h - y\|_{L^p,[0,t]}^p \leq 2^{p-1}\|h\|_{L^p,[0,t]}^p + 2^{p-1}L^p \int_0^t s^{p-1}\|y_h - y\|_{L^p,[0,s]}^p ds.$$

Assume $\delta = 1$. Then

$$(2.4) \quad |y_h - y|_{W^{1,p};[0,t]}^p \leq 2^{p-1}|h|_{W^{1,p};[0,t]}^p + 2^{p-1}L^p \int_0^t |y_h(s) - y(s)|^p ds$$

$$(2.5) \quad \leq 2^{p-1}|h|_{W^{1,p};[0,t]}^p + 2^{p-1}L^p \int_0^t \|y_h - y\|_{1-\text{var};[0,s]}^p ds$$

$$(2.6) \quad \leq 2^{p-1}|h|_{W^{1,p};[0,t]}^p + 2^{p-1}L^p \int_0^t s^{p-1}|y_h - y|_{W^{1,p};[0,s]}^p ds$$

where we used the embedding from [FV06, Theorem 1] in the last estimate. For $\delta \in (0, 1)$,

$$\begin{aligned} |y_h - y|_{W^{\delta,p};[0,t]}^p &\leq 2^{p-1}|h|_{W^{\delta,p};[0,t]}^p + 2^{p-1}L^p \iint_{[0,t]^2} \frac{\left(\int_u^v |y_h(s) - y(s)| ds\right)^p}{|v - u|^{1+\delta p}} du dv \\ &\leq 2^{p-1}|h|_{W^{\delta,p};[0,t]}^p + 2^{p-1}L^p \iint_{[0,t]^2} \frac{1}{|v - u|^{1+\delta p - p/q'}} du dv \left(\int_0^t |y_h(s) - y(s)|^q ds\right)^{\frac{p}{q}} \end{aligned}$$

where we used Hölder's inequality with $\frac{1}{q} + \frac{1}{q'} = 1$. Choosing q large enough such that $\delta < 1/q'$ ensures that the double integral is finite and we obtain

$$\begin{aligned} |y_h - y|_{W^{\delta,p};[0,t]}^p &\leq 2^{p-1}|h|_{W^{\delta,p};[0,t]}^p + C \left(\int_0^t \|y_h - y\|_{1/\delta-\text{var};[0,s]}^q ds\right)^{\frac{p}{q}} \\ &\leq 2^{p-1}|h|_{W^{\delta,p};[0,t]}^p + C \left(\int_0^t s^{\delta q - q/p} |y_h - y|_{W^{\delta,p};[0,s]}^q ds\right)^{\frac{p}{q}} \end{aligned}$$

using [FV06, Theorem 2] in the second estimate, thus

$$(2.7) \quad |y_h - y|_{W^{\delta,p};[0,t]}^q \leq C|h|_{W^{\delta,p};[0,t]}^q + C \int_0^t s^{\delta q - q/p} |y_h - y|_{W^{\delta,p};[0,s]}^q ds$$

by making the constant C larger if necessary. Combining the estimates (2.3), (2.6) and (2.7) with Gronwall's Lemma gives the claim. \square

Now let γ be the law of X on $C_0 = C_0([0, T], \mathbb{R}^d)$. As usual, \mathcal{H} denotes the corresponding Cameron–Martin space and we assume that \mathcal{H} is continuously embedded in C_0 . Set $C_\xi = C_\xi([0, T]; \mathbb{R}^d)$.

Theorem 2.3. *Let Y be the solution to the SDE (2.1) and let μ be the law of Y . Assume that b is Lipschitz continuous with Lipschitz constant L .*

(i) *If there is a continuous embedding*

$$(2.8) \quad \iota: \mathcal{H} \hookrightarrow W^{\delta,p}$$

for some $\delta \in (0, 1]$ and $p \in (1/\delta, \infty)$, then there is a constant $C = C(\delta, p, L)$ such that for every $\nu \in P(C_\xi)$

$$(2.9) \quad \inf_{\pi \in \Pi(\nu, \mu)} \int_{C_\xi \times C_\xi} d_{W^{\delta,p}}(x, y)^2 d\pi(x, y) \leq C \|\iota\|_{\mathcal{H} \hookrightarrow W^{\delta,p}}^2 H(\nu | \mu).$$

(ii) *If there is a continuous embedding*

$$(2.10) \quad \iota: \mathcal{H} \hookrightarrow C^{q-var}$$

for some $q \in [1, \infty)$, then for every $\nu \in P(C_\xi)$

$$(2.11) \quad \inf_{\pi \in \Pi(\nu, \mu)} \int_{C_\xi \times C_\xi} d_{q-var}(x, y)^2 d\pi(x, y) \leq 2e^{2LT} \|\iota\|_{\mathcal{H} \hookrightarrow C^{q-var}}^2 H(\nu | \mu).$$

Proof. Follows from Theorem 1.5, Lemma 4.3 and Lemma 2.2 resp. Lemma 2.1. \square

Remark 2.4. *Embeddings of the form (2.10) play a crucial role in Gaussian rough paths theory and we will revisit them also in the next section. Sufficient conditions for such embeddings are given in [FGGR13].*

Example 2.5. *We consider several Gaussian processes as driving signals for (2.1):*

- (1) *In the case of a Brownian motion, $\mathcal{H} = W_0^{1,2}$ and (2.9) holds. This was already shown in [DGW04, Proposition 5.4].*
- (2) *Let $X = B^H$ be a fractional Brownian motion with Hurst parameter $H \in (0, 1/2)$. Then it is known (cf. [FV06, Theorem 3]) that \mathcal{H} is compactly embedded in $W^{\delta,2}$ for every $\delta < H + 1/2$, thus (2.9) holds for every $\delta \in (1/2, H + 1/2)$. In [FGGR13, Theorem 1 and Example 2.7] it is shown that (2.10) holds with the choice $q = \frac{1}{H+1/2}$, hence we may conclude (2.11).*
- (3) *Many more examples for (2.11) (driving signals might be bifractional Brownian motion, Volterra processes, random Fourier series...) may be derived from [FGGR13, Theorem 1 and Examples 2.3 – 2.16].*

Note that in all these cases, the (fractional) Sobolev norms of the solution paths to (2.1) are in general not finite resp. they will not have finite q -variation almost surely.

2.2. SDEs with multiplicative noise. Next we will consider SDEs with multiplicative noise. Let $Y: [0, T] \rightarrow \mathbb{R}^m$ be the solution to

$$(2.12) \quad dY_t = f_0(Y_t) dt + \sum_{i=1}^d f_i(Y_t) \circ dW_t^i, \quad Y_0 = \xi \in \mathbb{R}^m$$

where $W = (W^1, \dots, W^d)$ is a d -dimensional Brownian motion and $f_0, f_1, \dots, f_d: \mathbb{R}^m \rightarrow \mathbb{R}^m$. In contrast to the additive noise case, the solution map $I_f(\cdot, \xi): C_0([0, T], \mathbb{R}^d) \rightarrow C_\xi([0, T], \mathbb{R}^m)$ which assigns to each Brownian path the solution path to the SDE (2.12) will in general *not* be

(Lipschitz-) continuous. This issue can be overcome using Lyons' rough paths theory. Indeed, rough paths theory shows that there is a Polish space $\mathcal{D}_g^{0,p}$ such that the diagram

$$(2.13) \quad \begin{array}{ccc} & \mathcal{D}_g^{0,p} & \\ & \uparrow S & \searrow \mathbf{I}_f(\cdot, \xi) \\ C_0 & \xrightarrow{I_f(\cdot, \xi)} & C_\xi \end{array}$$

commutes almost surely and the map $\mathbf{I}_f(\cdot, \xi): \mathcal{D}_g^{0,p} \rightarrow C_\xi$ is locally Lipschitz continuous. The map $S: C_0 \rightarrow \mathcal{D}_g^{0,p}$ is constructed w.r.t. the Wiener measure on the path space C_0 . Using a pathwise approach, one is not restricted to Wiener measure and it is indeed possible to construct lift maps S_γ w.r.t. more general Gaussian measures γ (cf. [CQ02], [FV10a]). In this case, one defines $I_f(\cdot, \xi) := \mathbf{I}_f(S(\cdot), \xi)$ which gives rise to solutions of SDEs of the form

$$(2.14) \quad dY_t = \sum_{i=1}^d f_i(Y_t) \circ dX_t^i, \quad Y_0 = \xi \in \mathbb{R}^m$$

where $X = (X^1, \dots, X^d)$ is a Gaussian process (for simplicity, we dropped the drift term here which does not cause problems anyway). Our key result will be that for *Brownian-like* Gaussian processes (we will be more precise later), we have an estimate of the form

$$\|I_f(x, \xi) - I_f(y, \xi)\|_{p\text{-var}} \leq L(y) d_{\mathcal{H}}(x, y)$$

almost surely for every $x, y \in C_0$ where L is a random variable which possesses every moment w.r.t. the Gaussian measure γ . Together with Lemma 4.3, this immediately yields our main result which is stated in Theorem 2.11.

We will not make an attempt to give an overview of rough paths theory since we will use it merely as a tool. Instead, we refer to the monographs [LQ02], [LCL07], [FV10b] and [FH]. The terms and notation we are using coincides with the one from [FV10b] with the addition that we use the symbol $\mathcal{D}_g^{0,p}$ for the geometric p -variation rough paths space $C_0^{0,p\text{-var}}([0, T]; G^{[p]}(\mathbb{R}^d))$.

We start with some deterministic estimates for rough paths. If ω is a control function and $\alpha > 0$, recall the definition of $N_\alpha(\omega; [s, t])$ resp. of $N_\alpha(\mathbf{x}; [s, t])$ for geometric rough paths \mathbf{x} ([CLL13], [FR13]). The next proposition is a version of [BFRS13, Theorem 4] for the p -variation metric.

Proposition 2.6. *Let \mathbf{x}^1 and \mathbf{x}^2 be weakly geometric p -rough paths for some $p \geq 1$. Consider the rough differential equations (RDEs)*

$$dy_t^j = f^j(y_t^j) d\mathbf{x}_t^j; \quad y_S^j \in \mathbb{R}^m$$

for $j = 1, 2$ on some interval $[S, T]$ where $f^1 = (f_i^1)_{i=1, \dots, m}$ and $f^2 = (f_i^2)_{i=1, \dots, m}$ are two families of vector fields, $\theta > p$ and β is a bound on² $|f^1|_{Lip^\theta}$ and $|f^2|_{Lip^\theta}$. Then for every $\alpha > 0$ there is a

²We mean Lipschitz in the sense of Stein, cf. [FV10b, Chapter 10]

constant $C = C(\theta, p, \beta, \alpha)$ such that

$$\begin{aligned} d_{p-\text{var};[S,T]}(y^1, y^2) &\leq C \left[|y_S^1 - y_S^2| + |f^1 - f^2|_{\text{Lip}^{\theta-1}} + \rho_{p-\text{var};[S,T]}(\mathbf{x}^1, \mathbf{x}^2) \right] \\ &\quad \times (\|\mathbf{x}^1\|_{p-\text{var};[S,T]} + \|\mathbf{x}^2\|_{p-\text{var};[S,T]} + 1) \\ &\quad \times \exp \left\{ C \left(N_\alpha(\mathbf{x}^1; [S, T]) + N_\alpha(\mathbf{x}^2; [S, T]) + 1 \right) \right\} \end{aligned}$$

holds.

Proof. The proof follows [BFRS13, Lemma 8 and Theorem 4]. Let ω be a control function such that $\sup_{s < t} \frac{\|\mathbf{x}^j\|}{\omega(s, t)^{1/p}} \leq 1$ for $j = 1, 2$. Set $\bar{y} := y^1 - y^2$ and

$$\kappa := \frac{|f^1 - f^2|_{\text{Lip}^{\theta-1}}}{\beta} + \rho_{p-\omega;[S,T]}(\mathbf{x}^1, \mathbf{x}^2).$$

We claim that there is a constant $C = C(\theta, p)$ such that for every $s < t$,

$$(2.15) \quad \|\bar{y}\|_{p-\text{var};[s,t]} \leq C\beta\omega(s, t)^{\frac{1}{p}} (\|\bar{y}\|_{\infty;[s,t]} + \kappa) \exp \{C\beta^p(N_\alpha(\omega; [s, t]) + 1)\}.$$

Indeed, as it was shown in the proof of [BFRS13, Lemma 8],

$$2^{1-p}|\bar{y}_{s,v}|^p \leq C\beta^p\omega(u, v)(|\bar{y}_u| + \kappa)^p \exp \{C\beta^p\omega(u, v)\} + |\bar{y}_{s,u}|^p$$

for every $[u, v] \subseteq [s, t]$. Thus if $s = \tau_0 < \dots < \tau_M < \tau_{M+1} = v$,

$$|\bar{y}_{s,v}|^p \leq 2^{(M+1)(p-1)} C\beta^p\omega(s, v)(\|\bar{y}\|_{\infty;[s,v]} + \kappa)^p \exp \left\{ C\beta^p \sum_{i=0}^M \omega(\tau_i, \tau_{i+1}) \right\}$$

for every $s \leq v \leq t$. Choosing $\tau_0 = s$, $\tau_{i+1} = \inf_t \{\omega(\tau_i, t) \geq \alpha\} \wedge v$ gives

$$|\bar{y}_{s,v}|^p \leq C\beta^p\omega(s, v)(\|\bar{y}\|_{\infty;[s,v]} + \kappa)^p \exp \{C\beta^p(N_\alpha(\omega; [s, v]) + 1)\}$$

for every $s \leq v \leq t$ and (2.15) follows. Now we can use the conclusion from [BFRS13, Lemma 8] to see that

$$\|\bar{y}\|_{p-\text{var};[s,t]} \leq C\beta\omega(s, t)^{\frac{1}{p}} (\|\bar{y}_s\| + \kappa) \exp \{C\beta^p(N_\alpha(\omega; [s, t]) + 1)\}$$

holds for every $s < t$. We conclude as in [BFRS13, Theorem 4]. \square

Lemma 2.7. *Let \mathbf{x} be a weakly geometric p -rough path with $p \in [2, 3)$ and h a path of finite q -variation with $1 \leq q \leq p$ and $\frac{1}{p} + \frac{1}{q} > 1$. Then there is a constant $C = C(p, q)$ such that*

$$\rho_{p-\text{var};[S,T]}(T_h(\mathbf{x}), \mathbf{x}) \leq C_{p,q}(1 \vee \|x\|_{p-\text{var};[S,T]})(\|h\|_{q-\text{var};[S,T]} + \|h\|_{q-\text{var};[S,T]}^2).$$

Proof. Recall that

$$\rho_{p-\text{var};[S,T]}(\mathbf{x}, \mathbf{y}) = \sup_{D \in \mathcal{P}([S, T])} \left(\sum_{t_i \in D} |x_{t_i, t_{i+1}} - y_{t_i, t_{i+1}}|^p \right)^{\frac{1}{p}} + \sup_{D \in \mathcal{P}([S, T])} \left(\sum_{t_i \in D} |\mathbf{x}_{t_i, t_{i+1}}^2 - \mathbf{y}_{t_i, t_{i+1}}^2|^{p/2} \right)^{\frac{2}{p}}.$$

Therefore, we immediately obtain

$$\rho_{p-\text{var};[S,T]}(T_h(\mathbf{x}), \mathbf{x}) \leq \|h\|_{q-\text{var};[S,T]} + \sup_{D \in \mathcal{P}([S, T])} \left(\sum_{t_i \in D} |T_h(\mathbf{x})_{t_i, t_{i+1}}^2 - \mathbf{x}_{t_i, t_{i+1}}^2|^{p/2} \right)^{\frac{2}{p}}.$$

Concerning the second term, fix some $D \in \mathcal{P}([S, T])$. We have

$$\sum_{t_i \in D} |T_h(\mathbf{x})_{t_i, t_{i+1}}^2 - \mathbf{x}_{t_i, t_{i+1}}^2|^{p/2} = \sum_{t_i \in D} \left| \int_{\Delta_{t_i, t_{i+1}}^2} d(x+h) \otimes d(x+h) - \int_{\Delta_{t_i, t_{i+1}}^2} dx \otimes dx \right|^{p/2}$$

and

$$\begin{aligned} & \left| \int_{\Delta_{t_i, t_{i+1}}^2} d(x+h) \otimes d(x+h) - \int_{\Delta_{t_i, t_{i+1}}^2} dx \otimes dx \right| \\ & \leq \left| \int_{\Delta_{t_i, t_{i+1}}^2} dh \otimes d(x+h) \right| + \left| \int_{\Delta_{t_i, t_{i+1}}^2} dx \otimes dh \right| \\ & \leq C_{p,q} \|h\|_{q\text{-var};[t_i, t_{i+1}]} (\|x+h\|_{p\text{-var};[t_i, t_{i+1}]} + \|x\|_{p\text{-var};[t_i, t_{i+1}]}) \end{aligned}$$

by the estimates for the Young integral. From Hölder's inequality,

$$\begin{aligned} \sum_{t_i \in D} |T_h(\mathbf{x})_{t_i, t_{i+1}}^2 - \mathbf{x}_{t_i, t_{i+1}}^2|^{p/2} & \leq C_{p,q} \left(\sum_{t_i} \|h\|_{q\text{-var};[t_i, t_{i+1}]}^q \right)^{\frac{p}{2q}} \left(\sum_{t_i} \|x+h\|_{p\text{-var};[t_i, t_{i+1}]}^p + \|x\|_{p\text{-var};[t_i, t_{i+1}]}^p \right)^{\frac{1}{2}} \\ & \leq C_{p,q} \|h\|_{q\text{-var};[S,T]}^{p/2} (\|x+h\|_{p\text{-var};[S,T]}^{p/2} + \|x\|_{p\text{-var};[S,T]}^{p/2}). \end{aligned}$$

and the result follows from the triangle inequality for the p -variation seminorm and standard estimates. \square

Lemma 2.8. *Let $\mathbf{x}^1 := \mathbf{x}$ and $\mathbf{x}^2 := T_h(\mathbf{x})$ where \mathbf{x} is a weakly geometric p -rough path for some $p \in [1, 3)$ and h is a path of finite q -variation with $\frac{1}{p} + \frac{1}{q} > 1$. Consider the solutions y^1 and y^2 to the RDEs as in Proposition 2.6 with $f^1 = f^2$ and $y_S^1 = y_S^2$. Then*

$$d_{p\text{-var};[S,T]}(y^1, y^2) \leq C \exp\{C(N_1(\mathbf{x}; [S, T]) + 1)\} (\|h\|_{q\text{-var};[S,T]} \vee \|h\|_{q\text{-var};[S,T]}^q)$$

where C is a constant depending on p, q, θ and β .

Proof. We will only consider the case $p \in [2, 3)$, the case $p \in [1, 2)$ is similar (and easier). Let $\|h\|_{q\text{-var};[S,T]} \leq 1$. We claim that

$$(2.16) \quad d_{p\text{-var};[S,T]}(y^1, y^2) \leq C \exp\{C(N_1(\mathbf{x}; [S, T]) + 1)\} \|h\|_{q\text{-var};[S,T]}$$

holds for some constant C . Indeed: From Proposition 2.6 we know that for every $\alpha > 0$,

$$\begin{aligned} d_{p\text{-var};[S,T]}(y^1, y^2) & \leq C(\|\mathbf{x}\|_{p\text{-var};[S,T]} + \|T_h(\mathbf{x})\|_{p\text{-var};[S,T]} + 1) \\ & \quad \times \exp\{C(N_\alpha(\mathbf{x}; [S, T]) + N_\alpha(T_h(\mathbf{x}); [S, T]) + 1)\} \rho_{p\text{-var};[S,T]}(\mathbf{x}, T_h(\mathbf{x})). \end{aligned}$$

Using Lemma 3.4, [FV10b, Theorem 9.33] and the assumption $\|h\|_{q\text{-var};[S,T]} \leq 1$ shows that

$$d_{p\text{-var};[S,T]}(y^1, y^2) \leq C(\|\mathbf{x}\|_{p\text{-var};[S,T]} + 1) \exp\{C(N_1(\mathbf{x}; [S, T]) + 1)\} \rho_{p\text{-var};[S,T]}(\mathbf{x}, T_h(\mathbf{x})).$$

for a larger constant C and α chosen appropriately. Applying Lemma 2.7 shows (2.16), using the estimate $\|\mathbf{x}\|_{p\text{-var};[S,T]} \leq N_1(\mathbf{x}; [S, T]) + 1$ which was proven in [FR13, Lemma 4].

Now let $\|h\|_{q\text{-var};[S,T]} \geq 1$. In this case,

$$\begin{aligned} d_{p\text{-var};[S,T]}(y^1, y^2) & \leq \|y^1\|_{p\text{-var};[S,T]} + \|y^2\|_{p\text{-var};[S,T]} \\ & \leq C(N_\alpha(\mathbf{x}; [S, T]) + N_\alpha(T_h(\mathbf{x}); [S, T]) + 1) \end{aligned}$$

using the deterministic estimates for the Itô–Lyons map proven in [FR13]. With Lemma 3.4, we conclude that

$$d_{p-\text{var};[S,T]}(y^1, y^2) \leq C(N_1(\mathbf{x}; [S, T]) + 1) \|h\|_{q-\text{var};[S,T]}^q$$

with α chosen appropriately and a larger constant C . □

We will now make further assumptions on our lift map S and on the Cameron–Martin space \mathcal{H} . Suppose that

(i) There is a continuous embedding

$$\iota: \mathcal{H} \hookrightarrow C^{q-\text{var}}([0, T], \mathbb{R}^d); \quad 1 \leq q \leq p$$

with $\frac{1}{p} + \frac{1}{q} > 1$ (note that this implies $1 \leq q < 2$ when $p \geq 2$).

(ii) The set

$$\{x \in C_0 \mid S(x + h) = T_h(S(x)) \text{ for all } h \in \mathcal{H}\} =: \tilde{C}_0$$

has full measure.

Remark 2.9. *Assumption (i) and (ii) are trivially fulfilled when $p \in [1, 2)$. Both assumptions are fulfilled for lifts in the sense of Friz–Victoir (cf. [FV10a], [FV10b, Proposition 15.7 and Lemma 15.58] and [FGGR13, Theorem 1]). In particular, they hold for the Stratonovich lift of the Brownian motion with $q = 1$.*

Under these two conditions, the following Proposition is an immediate consequence of Lemma 2.8.

Proposition 2.10. *Consider the RDEs as in Proposition 2.6 with $p \in [1, 3)$, $f^1 = f^2$, $y_0^1 = y_0^2$ and assume that $S: C_0 \rightarrow \mathcal{D}_g^{0,p}$ is a lift map such that the diagram (2.13) commutes. Then*

$$d_{p-\text{var}}(y^1, y^2) \leq L(x^1) (\|\iota\|_{\mathcal{H} \hookrightarrow C^{q-\text{var}}} \vee \|\iota\|_{\mathcal{H} \hookrightarrow C^{q-\text{var}}}^q) (d_{\mathcal{H}}(x^1, x^2) \vee d_{\mathcal{H}}(x^1, x^2)^q)$$

for all $x^1, x^2 \in C_0$ where

$$L(x) = C \exp \{C(N_1(S(x); [0, T]) + 1)\}$$

and C is a constant depending on p, q, θ and β .

The next theorem is our main result for the multiplicative case.

Theorem 2.11. *Let Y be the solution to the SDE (2.14) defined pathwise via the diagram (2.13) and let μ be the law of Y . Assume that $q = 1$. Then for every $\varepsilon > 0$ there is a constant C depending on ε, p, θ and β such that for every $\nu \in P(C_\xi)$,*

$$\inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{C_\xi \times C_\xi} d_{p-\text{var}}(x, y)^{2-\varepsilon} d\pi(x, y) \right)^{\frac{1}{2-\varepsilon}} \leq C \|\iota\|_{\mathcal{H} \hookrightarrow C^{1-\text{var}}} \sqrt{H(\nu \mid \mu)}.$$

Proof. From [CLL13, Theorem 6.3] we know that $N_1(S; [0, T])$ has Gaussian tails w.r.t. γ , hence $\|\exp\{C(N_1(S; [0, T]) + 1)\}\|_{L^q(\gamma)} < \infty$ for every $q \in [1, \infty)$. The assertion follows from Theorem 1.5, Proposition 2.10 and Lemma 4.3. □

- Remark 2.12.** (i) *The transportation–cost inequality in Theorem 2.11 holds in particular for the Stratonovich solution to a multiplicative SDE driven by a Brownian motion. As already mentioned in the introduction, this extends the known results in two ways. First, it is seen to hold for the (stronger) p -variation metric. Second, it holds for the parameter $2 - \varepsilon$, any $\varepsilon > 0$, without the assumption that Y is contracting in the L^2 -sense as in [WZ04].*
- (ii) *Many more Gaussian processes fulfill assumption (i) and (ii) for $q = 1$ and therefore may be considered as driving signals in Theorem 2.11, e.g. fractional Brownian motions with Hurst parameter $H \geq \frac{1}{2}$ (in which case we extend results from [Sau12]), Brownian bridges, Ornstein–Uhlenbeck processes, bifractional Brownian motions and random Fourier series; see [FGGR13] for a detailed account and even more examples.*

3. TAIL ESTIMATES FOR FUNCTIONALS

It is well known that transportation–cost inequalities imply Gaussian measure concentration. This was first discovered by Marton ([Mar86], [Mar96]). In the following, we show how to modify her argument in order to deduce tail estimates in a more general context.

Theorem 3.1. *Let V be a linear Polish space and let μ be a probability measure defined on the Borel σ -algebra of V . Assume that there is a normed space $\mathcal{B} \subseteq V$. Let $d_{\mathcal{B}}: V \times V \rightarrow \mathbb{R} \cup \{+\infty\}$ be defined as*

$$d_{\mathcal{B}}(x, y) = \begin{cases} \|x - y\|_{\mathcal{B}} & \text{if } x - y \in \mathcal{B} \\ +\infty & \text{otherwise} \end{cases}$$

and assume that there is a $p \in [1, \infty)$ and a constant C such that for every $\nu \in P(V)$,

$$\inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{V \times V} d_{\mathcal{B}}(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}} \leq C \sqrt{H(\nu | \mu)}.$$

Let (E, d) be some metric space, $f: V \rightarrow E$ be measurable w.r.t. the Borel σ -algebra and assume that there is an $r_0 \geq 0$ and some element $e \in E$ such that

$$\mu \{x \in V : d(f(x), e) \leq r_0\} =: a > 0.$$

- (i) *Assume that there is a nullset \mathcal{N} such that*

$$d(f(x + h), f(x)) \leq c(x) \|h\|_{\mathcal{B}}$$

holds for every x outside \mathcal{N} and $h \in \mathcal{B}$ where $c \in L^q(\mu)$ with $q \in (1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\mu \{x \in V : d(f(x), e) > r\} \leq \exp \left\{ - \left(\frac{r - r_1}{C \|c\|_{L^q(\mu)}} \right)^2 \right\}$$

for all $r \geq r_1$ where $r_1 = r_0 + \|c\|_{L^q(\mu)} \sqrt{2 \log(a^{-1})}$.

- (ii) *Assume that there is a nullset \mathcal{N} such that*

$$d(f(x + h), e) \leq c(x) (g(x) + \|h\|_{\mathcal{B}})$$

holds for every x outside \mathcal{N} and $h \in \mathcal{B}$ with c as in (i) and $\langle c, g \rangle = \int c g d\mu < \infty$. Then

$$\mu \{x \in V : d(f(x), e) > r\} \leq \exp \left\{ - \left(\frac{r - r_1}{C \|c\|_{L^q(\mu)}} \right)^2 \right\}$$

for all $r \geq r_2$ where $r_2 = r_0 + 4\langle c, g \rangle + \|c\|_{L^q(\mu)} \sqrt{2 \log(a^{-1})}$.

In particular, in both cases, $d(f(\cdot), e): V \rightarrow \mathbb{R}$ has Gaussian tails.

Proof. For $x, y \in V$ set

$$d_f(x, y) = d(f(x), f(y)).$$

For any measurable set $A \subseteq V$ and $r \geq 0$ we define

$$A^r := \{x \in V : \text{there is an } \bar{x} \in A \text{ such that } d_f(x, \bar{x}) \leq r\}.$$

Fix some $r \geq 0$ and set $B := (A^r)^c$. Assume first that A and B have positive measure. On V , we define the measures

$$d\mu_A := \frac{\mathbb{1}_A}{\mu(A)} d\mu \quad \text{and} \quad d\mu_B := \frac{\mathbb{1}_B}{\mu(B)} d\mu.$$

Then

$$\begin{aligned} r &\leq \inf_{\pi \in \Pi(\mu_A, \mu_B)} \int_{V \times V} d_f(x, y) d\pi(x, y) \\ &\leq \inf_{\pi \in \Pi(\mu_A, \mu)} \int_{V \times V} d_f(x, y) d\pi(x, y) + \inf_{\pi \in \Pi(\mu_B, \mu)} \int_{V \times V} d_f(x, y) d\pi(x, y) \end{aligned}$$

where we used symmetry and the triangle inequality which can be deduced from Lemma 4.2. Set $\tilde{V} := V \setminus \mathcal{N}$ and let $x, y \in \tilde{V}$. If $x - y = h \in \mathcal{B}$, we have

$$(3.1) \quad d_f(x, y) = d(f(y + h), f(y)) \leq c(y) \|h\|_{\mathcal{B}}$$

in case (i) and

$$d_f(x, y) \leq 2c(y)g(y) + c(y)\|h\|_{\mathcal{B}}$$

in case (ii). In the following, we will only show the conclusion stated in (i), part (ii) is similar. From (3.1), we can deduce that for all $x, y \in \tilde{V}$,

$$d_f(x, y) \leq c(y)d_{\mathcal{B}}(x, y).$$

Since \tilde{V} has full measure for μ , it follows that this inequality holds π almost surely for every $\pi \in \Pi(\mu_A, \mu)$ resp. $\pi \in \Pi(\mu_B, \mu)$, hence

$$\begin{aligned} r &\leq \inf_{\pi \in \Pi(\mu_A, \mu)} \int_{V \times V} c(y)d_{\mathcal{B}}(x, y) d\pi(x, y) + \inf_{\pi \in \Pi(\mu_B, \mu)} \int_{V \times V} c(y)d_{\mathcal{B}}(x, y) d\pi(x, y) \\ &\leq \|c\|_{L^q} \inf_{\pi \in \Pi(\mu_A, \mu)} \left(\int_{V \times V} d_{\mathcal{B}}(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}} + \|c\|_{L^q} \inf_{\pi \in \Pi(\mu_B, \mu)} \left(\int_{V \times V} d_{\mathcal{B}}(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}} \\ &\leq \|c\|_{L^q} C \sqrt{H(\mu_A | \mu)} + \|c\|_{L^q} C \sqrt{H(\mu_B | \mu)} \\ &= \|c\|_{L^q} C \sqrt{\log(\mu(A)^{-1})} + \|c\|_{L^q} C \sqrt{\log(\mu(B)^{-1})}. \end{aligned}$$

Rearranging terms, we see that

$$1 - \exp \left\{ - \left(\frac{r - \tilde{r}}{C\|c\|_{L^q}} \right)^2 \right\} \leq \mu(A^r)$$

for every $r \geq \tilde{r}$ where $\tilde{r} := C\|c\|_{L^q} \sqrt{\log(\mu(A)^{-1})}$. Now set

$$A := \{x \in V : d(f(x), e) \leq r_0\}.$$

Then we have for every $r \geq 0$,

$$A^r \subseteq \{x \in V : d(f(x), e) \leq r_0 + r\}.$$

If $\mu(B) = 0$, it follows that $\{x \in V : d(f(x), e) \leq r_0 + r\}$ has full measure. In other words, $d(f(\cdot), e)$ is bounded almost surely and the claimed estimate is trivial. If $\mu(B) > 0$, we can use our calculations above to conclude that

$$1 - \exp \left\{ - \left(\frac{r - \tilde{r}}{C \|c\|_{L^q}} \right)^2 \right\} \leq \mu\{x \in V : d(f(x), e) \leq r_0 + r\}$$

holds for every $r \geq \tilde{r}$ and the claim follows. \square

Remark 3.2. *In the Gaussian case, the assumptions from Theorem 3.1 hold with V being a Banach space, $\mathcal{B} = \mathcal{H}$ being continuously embedded in V and $p = q = 2$. It is well known (cf. [Bog98, 4.5.6. Theorem]) that \mathcal{H} -Lipschitzian functions have Gaussian tails. Part (i) in Theorem 3.1 shows that this assumption can be relaxed by assuming that we can control the Lipschitz constant as in (3.1).*

Example 3.3. (1) *In case of a Gaussian Banach space (E, \mathcal{H}, γ) , choosing $f = \|\cdot\|_E$ gives the usual Fernique theorem.*

(2) *In case of existence of a lift map $S: C_0 \rightarrow \mathcal{D}_g^{0,p}$ to a rough paths space as in Section 2.2, we obtain the Fernique estimate for Gaussian rough paths (see [FO10]) by setting $f(x) = d_{p\text{-var}}(S(x), e)$.*

(3) *Choosing $f(x) = \|I_f(x, \xi)\|_{p\text{-var}}$ and using Proposition 2.10 together with the tail estimates for the counting process $N_1(S(x); [0, T])$ (cf. [CLL13] and the forthcoming example 3.5), we see that the p -variation norm of solutions to rough differential equations with sufficiently smooth driving signals (i.e. Gaussian processes with Cameron–Martin paths of bounded variation) have Gaussian tails (this result is however not new and was already obtained in [FR13]).*

We want to show now how to deduce tail estimates for the counting process $N_1(S(X); [0, T])$ where X is a stochastic process for which we assume that its law satisfies a transportation–cost inequality on a path space. In the Gaussian case, these estimates can be seen as one of the key results in [CLL13] and we already used them several times in this work.

Lemma 3.4. *Let \mathbf{x} be a weakly geometric p -rough path and h be a path of bounded q -variation where $1 \leq q \leq p$ and $\frac{1}{p} + \frac{1}{q} > 1$. Then there is an $\alpha = \alpha(p, q)$ such that*

$$N_\alpha(T_h(\mathbf{x}); [0, T]) \leq (\|\mathbf{x}\|_{p\text{-var}}^p \vee (2N_1(\mathbf{x}; [0, T]) + 1)) + \|h\|_{q\text{-var}}^q.$$

Proof. We have

$$\begin{aligned} N_\alpha(\|T_h \mathbf{x}\|_{p\text{-var}}^p; [0, T]) &\leq N_\alpha(C_{p,q}(\|\mathbf{x}\|_{p\text{-var}}^p + \|h\|_{q\text{-var}}^p); [0, T]) \\ &= N_1(\|\mathbf{x}\|_{p\text{-var}}^p + \|h\|_{q\text{-var}}^p; [0, T]) \end{aligned}$$

with the choice $\alpha = C_{p,q}$, using [FV10b, Theorem 9.33]. By definition,

$$N_1(\|\mathbf{x}\|_{p\text{-var}}^p + \|h\|_{q\text{-var}}^p; [0, T]) \leq \sum_{\tau_i} \|\mathbf{x}\|_{p\text{-var}; [\tau_i, \tau_{i+1}]}^p + \|h\|_{q\text{-var}; [\tau_i, \tau_{i+1}]}^p$$

where (τ_i) is a finite partition of $[0, T]$ for which $\|\mathbf{x}\|_{p\text{-var};[\tau_i, \tau_{i+1}]}^p + \|h\|_{q\text{-var};[\tau_i, \tau_{i+1}]}^p \leq 1$ for every τ_i , and in particular $\|h\|_{q\text{-var};[\tau_i, \tau_{i+1}]}^p \leq \|h\|_{q\text{-var};[\tau_i, \tau_{i+1}]}^q$. Hence

$$\begin{aligned} N_1(\|\mathbf{x}\|_{p\text{-var}}^p + \|h\|_{q\text{-var}}^p; [0, T]) &\leq \sum_{\tau_i} \|\mathbf{x}\|_{p\text{-var};[\tau_i, \tau_{i+1}]}^p + \|h\|_{q\text{-var};[\tau_i, \tau_{i+1}]}^q \\ &\leq \sup_{\substack{(\tau_i) \in \mathcal{P}([0, T]) \\ \|\mathbf{x}\|_{p\text{-var};[\tau_i, \tau_{i+1}]} \leq 1}} \sum_{\tau_i} \|\mathbf{x}\|_{p\text{-var};[\tau_i, \tau_{i+1}]}^p + \|h\|_{q\text{-var};[0, T]}^q. \end{aligned}$$

The claim follows from [CLL13, Proposition 4.11]. \square

Example 3.5. Let X be a Gaussian process and consider the counting process

$$t \mapsto N_\alpha(\|S(X)\|_{p\text{-var}}^p; [0, t]) = N_\alpha(S(X); [0, t])$$

where S is a lift map as in Section 2.2. Lemma 3.4 shows that

$$N_\alpha(S(x+h); [0, t]) \leq \|S(x)\|_{p\text{-var};[0, t]}^p + \|\iota\|_{\mathcal{H} \hookrightarrow C^{q\text{-var}}}^q \|h\|_{\mathcal{H}}^q$$

almost surely, which implies that $N_\alpha(S(X); [0, t])^{\frac{1}{q}}$ has Gaussian tails for every $t \geq 0$, one of the main results obtained in [CLL13]. This implies in particular that the random variable L in Proposition 2.10 has moments of any order.

4. APPENDIX

4.1. Optimal transport and the Wasserstein metric. We start with an approximation result which will be used for the Wasserstein metric. The proof is taken from [Vil03, part 3 in the proof of Theorem 1.3].

Lemma 4.1. Let X and Y be Polish spaces and let μ and ν be probability measures on X resp. Y . Let $c_n: X \times Y \rightarrow \mathbb{R}_+$ be a nondecreasing sequence of bounded, continuous functions with $c_n \nearrow c$ pointwise where $c: X \times Y \rightarrow \mathbb{R}_+ \cup \{+\infty\}$. Then

$$\lim_{n \rightarrow \infty} \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c_n(x, y) d\pi(x, y) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\pi(x, y).$$

Proof. First, it can be shown (cf. [Vil03, p. 32]) that $\Pi(\mu, \nu)$ is compact in the weak topology. For $\pi \in \Pi(\mu, \nu)$, set

$$I_n(\pi) := \int_{X \times Y} c_n(x, y) d\pi(x, y) \quad \text{and} \quad I(\pi) = \int_{X \times Y} c(x, y) d\pi(x, y).$$

Since the c_n are nondecreasing, $\inf_{\pi \in \Pi(\mu, \nu)} I_n(\pi)$ is nondecreasing and

$$\inf_{\pi \in \Pi(\mu, \nu)} I_n(\pi) \leq \inf_{\pi \in \Pi(\mu, \nu)} I(\pi).$$

Therefore, the limit exists and it is enough to show that

$$\lim_{n \rightarrow \infty} \inf_{\pi \in \Pi(\mu, \nu)} I_n(\pi) \geq \inf_{\pi \in \Pi(\mu, \nu)} I(\pi).$$

From continuity of the c_n , we know that the infima are attained (cf. [Vil03, Theorem 1.3]), thus there are $\pi_n \in \Pi(\mu, \nu)$ such that

$$I_n(\pi_n) = \inf_{\pi \in \Pi(\mu, \nu)} I_n(\pi).$$

From compactness, there is a subsequence (π_{n_k}) and a $\pi_* \in \Pi(\mu, \nu)$ such that $\pi_{n_k} \rightarrow \pi_*$ weakly for $k \rightarrow \infty$. Now, whenever $n \geq m$, $I_n(\pi_n) \geq I_m(\pi_n)$ and

$$\lim_{n \rightarrow \infty} I_n(\pi_n) \geq \limsup_{n \rightarrow \infty} I_m(\pi_n) \geq I_m(\pi_*).$$

Monotone convergence gives

$$\lim_{m \rightarrow \infty} I_m(\pi_*) = I(\pi_*)$$

and thus

$$\lim_{n \rightarrow \infty} I_n(\pi_n) \geq \lim_{m \rightarrow \infty} I_m(\pi_*) = I(\pi_*) \geq \inf_{\pi \in \Pi(\mu, \nu)} I(\pi).$$

□

It is well known that the Wasserstein metric \mathcal{W}_p^d is indeed a metric when d is a metric. The next Lemma analyzes the respective properties more carefully.

Lemma 4.2. *Let X be Polish, $c: X \times X \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ be measurable and $p \geq 1$.*

- (i) *If $c(x, x) = 0$ for all $x \in X$, we have $\mathcal{W}_p^c(\nu, \nu) = 0$ for all $\nu \in P(X)$.*
- (ii) *Assume in addition that c is lower semicontinuous. If $c(x, y) = 0$ implies $x = y$, also $\mathcal{W}_p^c(\nu, \mu) = 0$ implies $\nu = \mu$.*
- (iii) *If c is symmetric, also \mathcal{W}_p^c is symmetric.*
- (iv) *If the triangle inequality holds for c , it also holds for \mathcal{W}_p^c .*

In particular, if $c = d$ is a (pseudo-)metric, \mathcal{W}_p^d defines a (possibly infinite) (pseudo-)metric on the space $P(X)$ for all $p \geq 1$.

Proof. (i), (ii) and (iii) are easy and can be shown as in [Vil03, Theorem 7.3]. It remains to prove (iv). Let $\nu_1, \nu_2, \nu_3 \in P(X)$ and let $\varepsilon > 0$. Choose $\pi_{12} \in \Pi(\nu_1, \nu_2)$ and $\pi_{23} \in \Pi(\nu_2, \nu_3)$ such that

$$\left(\int c(x, y)^p d\pi_{12}(x, y) \right)^{\frac{1}{p}} \leq \mathcal{W}_p^c(\nu_1, \nu_2) + \varepsilon$$

and the same for π_{23} and $\mathcal{W}_p^c(\nu_2, \nu_3)$. From the Gluing Lemma (cf. [Vil03, Lemma 7.6]) there is a probability measure $\pi \in P(X \times X \times X)$ such that $\pi_{12} = \pi(\cdot, \cdot, X)$ and $\pi_{23} = \pi(X, \cdot, \cdot)$. Set $\pi_{13} = \pi(\cdot, X, \cdot)$. Then, using the triangle inequality for c ,

$$\begin{aligned} \mathcal{W}_p^c(\nu_1, \nu_3) &\leq \left(\int c(x, y)^p d\pi_{13}(x, y) \right)^{\frac{1}{p}} \\ &\leq \left(\int c(x, y)^p d\pi_{12}(x, y) \right)^{\frac{1}{p}} + \left(\int c(x, y)^p d\pi_{23}(x, y) \right)^{\frac{1}{p}} \\ &\leq \mathcal{W}_p^c(\nu_1, \nu_2) + \mathcal{W}_p^c(\nu_2, \nu_3) + 2\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this shows the claim.

□

The next Lemma is a generalization of [DGW04, Lemma 2.1].

Lemma 4.3. *Let (X, \mathcal{F}) be a measurable space on which regular conditional distributions exist and let $c: X \times X \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ be a measurable function. Assume that there is a measure $\mu \in P(X)$ such that*

$$\inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{X \times X} c(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}} \leq \sqrt{CH(\nu | \mu)}$$

holds for every $\nu \in P(X)$ where C is some constant and $p \in [1, \infty)$. Let (Y, \mathcal{G}) be another measurable space, $\tilde{c}: Y \times Y \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ be a measurable function and assume that there is a measurable function $\Psi: X \rightarrow Y$ for which

$$\tilde{c}(\Psi(x), \Psi(y)) \leq L(y)c(x, y)$$

holds for every $x, y \in X_0$ where $X_0 \subseteq X$ has full measure w.r.t. μ and $L: X \rightarrow \mathbb{R} \cup \{+\infty\}$ is another measurable function. Set $\tilde{\mu} := \mu \circ \Psi^{-1}$. Then for every $\tilde{p} \in [1, p]$,

$$\inf_{\tilde{\pi} \in \Pi(\tilde{\nu}, \tilde{\mu})} \left(\int_{Y \times Y} \tilde{c}(x, y)^{\tilde{p}} d\tilde{\pi}(x, y) \right)^{\frac{1}{\tilde{p}}} \leq \|L\|_{L^q(\mu)} \sqrt{CH(\tilde{\nu} | \tilde{\mu})}$$

holds for every $\tilde{\nu} \in P(Y)$ where $q \in (1, \infty]$ is chosen such that $\frac{1}{q} + \frac{1}{p} = \frac{1}{\tilde{p}}$.

Proof. W.l.o.g. we may assume $C = 1$. Let $\tilde{\nu} \in P(Y)$ and assume that $H(\tilde{\nu} | \tilde{\mu}) < \infty$. Choose $\nu \in P(X)$ such that $\tilde{\nu} = \nu \circ \Psi^{-1}$ and $\nu \ll \mu$ (note that there is at least one ν which fulfills this condition; e.g. $\nu_0(dx) := \frac{d\tilde{\nu}}{d\tilde{\mu}}(\Psi(x))\mu(dx)$). Then

$$\begin{aligned} \inf_{\tilde{\pi} \in \Pi(\tilde{\nu}, \tilde{\mu})} \int \tilde{c}(x, y)^{\tilde{p}} d\tilde{\pi}(x, y) &\leq \inf_{\pi \in \Pi(\nu, \mu)} \int_{Y \times Y} \tilde{c}(x, y)^{\tilde{p}} d(\pi \circ (\Psi \times \Psi)^{-1})(x, y) \\ &= \inf_{\pi \in \Pi(\nu, \mu)} \int_{X \times X} \tilde{c}(\Psi(x), \Psi(y))^{\tilde{p}} d\pi(x, y). \end{aligned}$$

Since $\nu \ll \mu$, $X_0 \times X_0$ has full measure for every $\pi \in \Pi(\nu, \mu)$, therefore

$$\begin{aligned} \inf_{\pi \in \Pi(\nu, \mu)} \int_{X \times X} \tilde{c}(\Psi(x), \Psi(y))^{\tilde{p}} d\pi(x, y) &\leq \inf_{\pi \in \Pi(\nu, \mu)} \int_{X \times X} (L(y)c(x, y))^{\tilde{p}} d\pi(x, y) \\ &\leq \|L\|_{L^q(\mu)}^{\tilde{p}} \inf_{\pi \in \Pi(\nu, \mu)} \left(\int_{X \times X} c(x, y)^p d\pi(x, y) \right)^{\frac{\tilde{p}}{p}} \end{aligned}$$

by Hölder's inequality. The assertion follows from the identity

$$(4.1) \quad H(\tilde{\nu} | \tilde{\mu}) = \inf \{ H(\nu | \mu) \mid \nu \in P(X) \text{ s.t. } \nu \circ \Psi^{-1} = \tilde{\nu} \}$$

which holds under the assumption that regular conditional distributions exist on (X, \mathcal{F}) , see [DGW04, Lemma 2.1]. \square

REFERENCES

- [BFRS13] Christian Bayer, Peter K. Friz, Sebastian Riedel, and John Schoenmakers. From rough path estimates to multilevel Monte Carlo. *arXiv:1305.5779*, 2013.
- [BG99] Sergey G. Bobkov and Friedrich Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.
- [Bog98] Vladimir I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.
- [Bor75] Christer Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30(2):207–216, 1975.
- [CLL13] Thomas Cass, Christian Litterer, and Terry J. Lyons. Integrability and tail estimates for Gaussian rough differential equations. *Ann. Probab.*, 41(4):3026–3050, 2013.

- [CQ02] Laure Coutin and Zhongmin Qian. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields*, 122(1):108–140, 2002.
- [DGW04] Hacene Djellout, Arnaud Guillin, and Li-ming Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B):2702–2732, 2004.
- [DOR14] Joscha Diehl, Harald Oberhauser, and Sebastian Riedel. A Levy–area between Brownian motion and rough paths with applications to robust non-linear filtering and RPDEs. *arXiv:1301.3799v2*, 2014.
- [Dud89] Richard M. Dudley. *Real analysis and probability*. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.
- [DZ98] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.
- [FGGR13] Peter K. Friz, Benjamin Gess, Archil Gulisashvili, and Sebastian Riedel. Jain–Monrad criterion for rough paths and applications. *arXiv:1307.3460*, 2013.
- [FH] Peter K. Friz and Martin Hairer. *A Short Course on Rough Paths*. Preprint.
- [FO10] Peter K. Friz and Harald Oberhauser. A generalized Fernique theorem and applications. *Proc. Amer. Math. Soc.*, 138(10):3679–3688, 2010.
- [FR13] Peter K. Friz and Sebastian Riedel. Integrability of (non-)linear rough differential equations and integrals. *Stoch. Anal. Appl.*, 31(2):336–358, 2013.
- [FÜ04] Denis Feyel and Ali S. Üstünel. Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space. *Probab. Theory Related Fields*, 128(3):347–385, 2004.
- [FV06] Peter K. Friz and Nicolas B. Victoir. A note on the notion of geometric rough paths. *Probab. Theory Related Fields*, 136(3):395–416, 2006.
- [FV10a] Peter K. Friz and Nicolas B. Victoir. Differential equations driven by Gaussian signals. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(2):369–413, 2010.
- [FV10b] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.
- [GL07] Nathael Gozlan and Christian Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields*, 139(1-2):235–283, 2007.
- [GL13] Nicola Gigli and Michel Ledoux. From log Sobolev to Talagrand: a quick proof. *Discrete Contin. Dyn. Syst.*, 33(5):1927–1935, 2013.
- [Goz09] Nathael Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.*, 37(6):2480–2498, 2009.
- [LCL07] Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.
- [Led96] Michel Ledoux. Isoperimetry and Gaussian analysis. In *Lectures on probability theory and statistics (Saint-Flour, 1994)*, volume 1648 of *Lecture Notes in Math.*, pages 165–294. Springer, Berlin, 1996.
- [Led01] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [LQ98] Terry J. Lyons and Zhongmin Qian. Flow of diffeomorphisms induced by a geometric multiplicative functional. *Probab. Theory Related Fields*, 112(1):91–119, 1998.
- [LQ02] Terry J. Lyons and Zhongmin Qian. *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2002. Oxford Science Publications.
- [Lyo98] Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.
- [Mar86] Katalin Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory*, 32(3):445–446, 1986.
- [Mar96] Katalin Marton. Bounding \bar{d} -distance by informational divergence: a method to prove measure concentration. *Ann. Probab.*, 24(2):857–866, 1996.
- [OV00] Felix Otto and Cédric Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.
- [Sau12] Bruno Saussereau. Transportation inequalities for stochastic differential equations driven by a fractional Brownian motion. *Bernoulli*, 18(1):1–23, 2012.

- [Tal96] Michel P. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.
- [Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [Vil09] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [Wan02] Feng-Yu Wang. Transportation cost inequalities on path spaces over Riemannian manifolds. *Illinois J. Math.*, 46(4):1197–1206, 2002.
- [WZ04] Li-ming Wu and Zheng-liang Zhang. Talagrand’s T_2 -transportation inequality w.r.t. a uniform metric for diffusions. *Acta Math. Appl. Sin. Engl. Ser.*, 20(3):357–364, 2004.

E-mail address: `riedel@math.tu-berlin.de`

SEBASTIAN RIEDEL, INSTITUT FÜR MATHEMATIK, TECHNISCHE UNIVERSITÄT BERLIN, GERMANY